

Crawling Code Review Data from Phabricator

Abstract

Modern code review is typically supported by software tools. Researchers use data tracked in these tools to study code review practices. Popular tools in open-source and closed-source projects are Gerrit and Phabricator. While researchers can rely on Gerrit code review crawlers like Gerry, there is no such crawler tool for Phabricator. In this thesis, we develop a Python crawler for crawling code review data from Phabricator instances using its REST API. The tool works producing minimal server load, supports incremental crawling and reproducible crawl runs, and stores complete and genuine review data. The new tool is applied to the Phabricator instances of the open source projects Blender, Haskell, FreeBSD, and LLVM. The resulting data sets are of high quality, [..], [...] and can be used by researchers. We also present a short overview of the gathered dataset and its properties.